

# CHƯƠNG 2

## THU THẬP VÀ TRÌNH BÀY DỮ LIỆU



# MỘT SỐ CÁCH LẤY MẪU THƯỜNG DÙNG

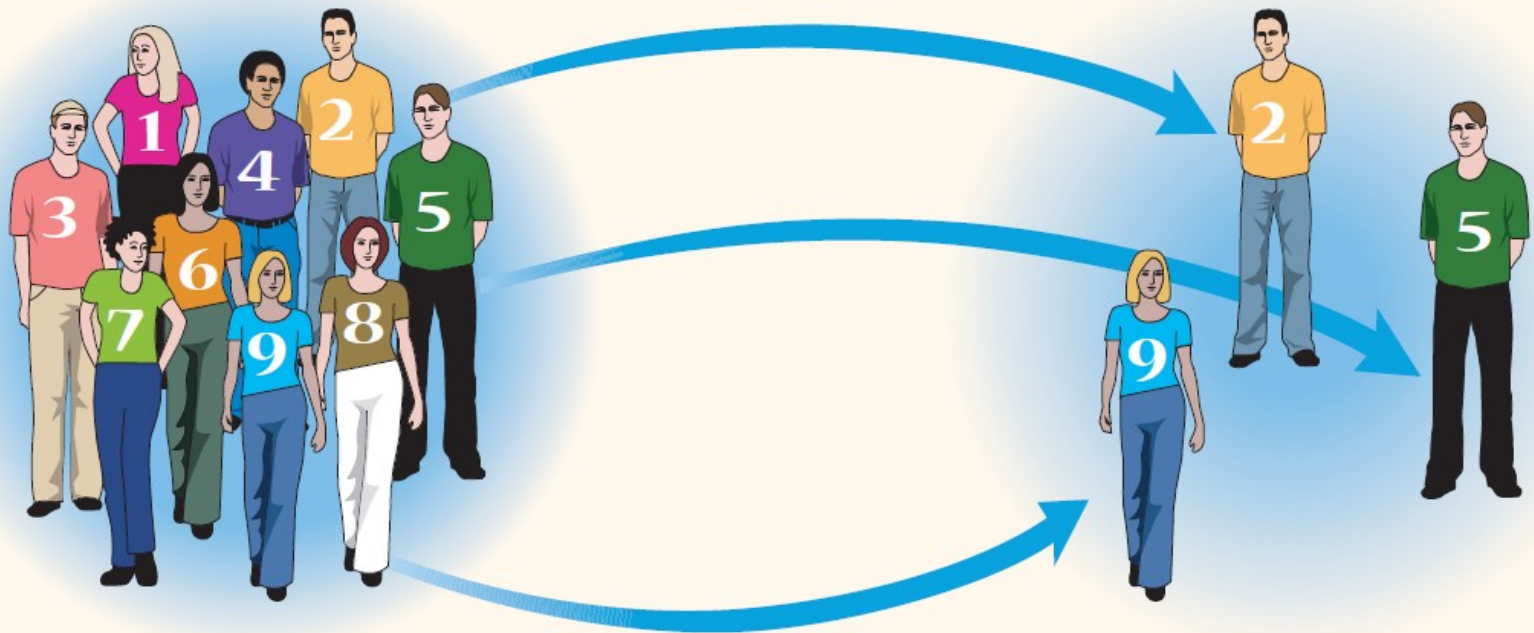
---

1. Chọn mẫu ngẫu nhiên.
2. Chọn mẫu phân tầng.
3. Chọn mẫu theo hệ thống.
4. Chọn mẫu theo khối.



# LẤY MẪU NGẪU NHIÊN

## Simple Random Sampling

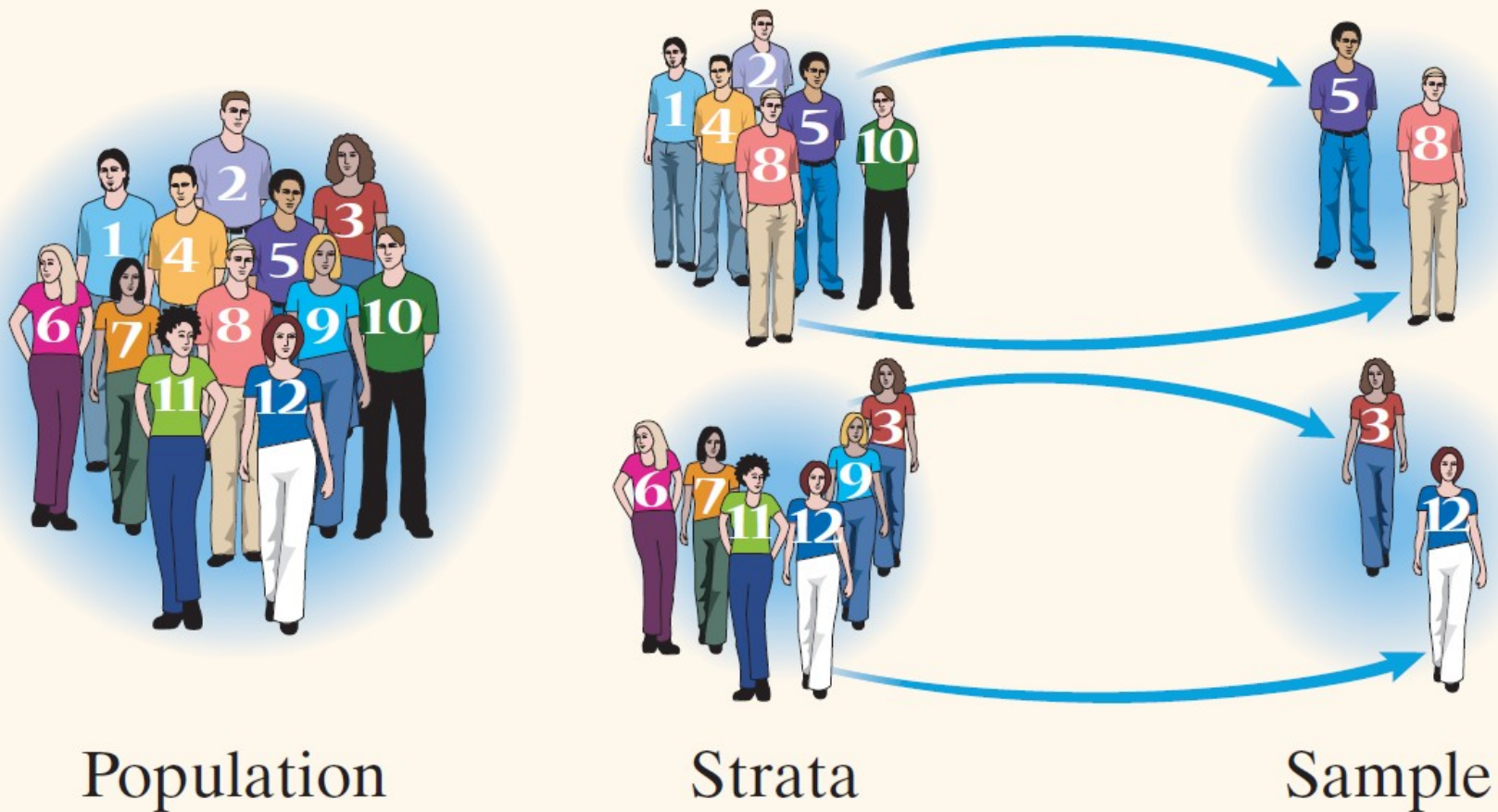


Population

Sample

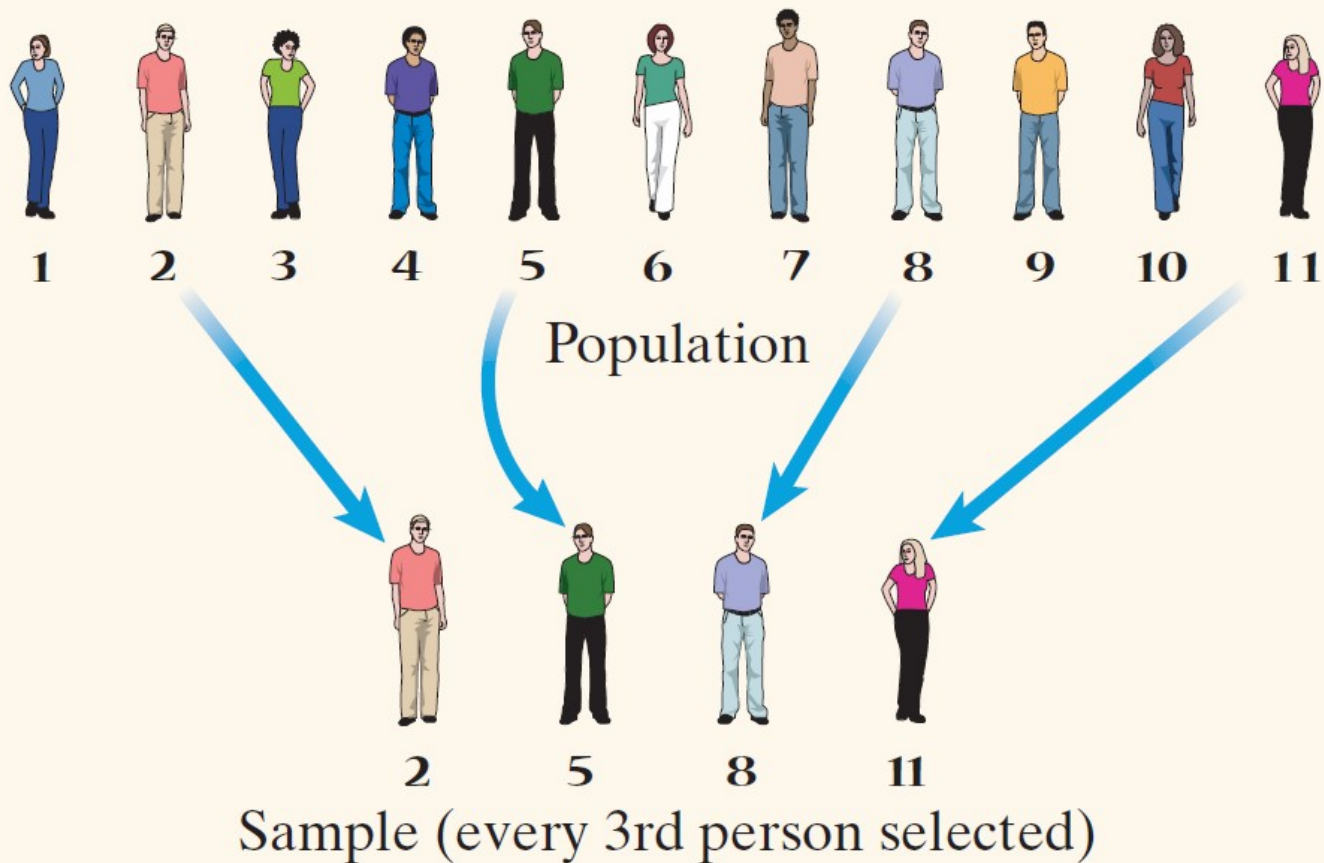
# LẤY MẪU PHÂN TẦNG

## Stratified Sampling



# LẤY MẪU HỆ THỐNG

























## Systematic Sampling





























# LẤY MẪU THEO KHỐI





## Cluster Sampling





					
1	2	5	6	9	10
					
3	4	7	8	11	12
					
13	14	17	18	21	22
					
15	16	19	20	23	24

Population

					
1	2	5	6	9	10
					
3	4	7	8	11	12
					
13	14	17	18	21	22
					
15	16	19	20	23	24

Cluster  
Population

	
1	2
	
3	4

	
17	18
	
19	20

Sample:  
Randomly  
Selected  
Clusters

# SẮP XẾP DỮ LIỆU ĐỊNH TÍNH

DATA FROM A SAMPLE OF 50 SOFT DRINK PURCHASES

Coca-Cola	Coca-Cola	Coca-Cola	Sprite	Coca-Cola
Diet Coke	Dr. Pepper	Diet Coke	Dr. Pepper	Diet Coke
Pepsi	Sprite	Coca-Cola	Pepsi	Pepsi
Diet Coke	Coca-Cola	Sprite	Diet Coke	Pepsi
Coca-Cola	Diet Coke	Pepsi	Pepsi	Pepsi
Coca-Cola	Coca-Cola	Coca-Cola	Coca-Cola	Pepsi
Dr. Pepper	Coca-Cola	Coca-Cola	Coca-Cola	Coca-Cola
Diet Coke	Sprite	Coca-Cola	Coca-Cola	Dr. Pepper
Pepsi	Coca-Cola	Pepsi	Pepsi	Pepsi
Pepsi	Diet Coke	Coca-Cola	Dr. Pepper	Sprite



# BẢNG PHÂN PHỐI TẦN SỐ

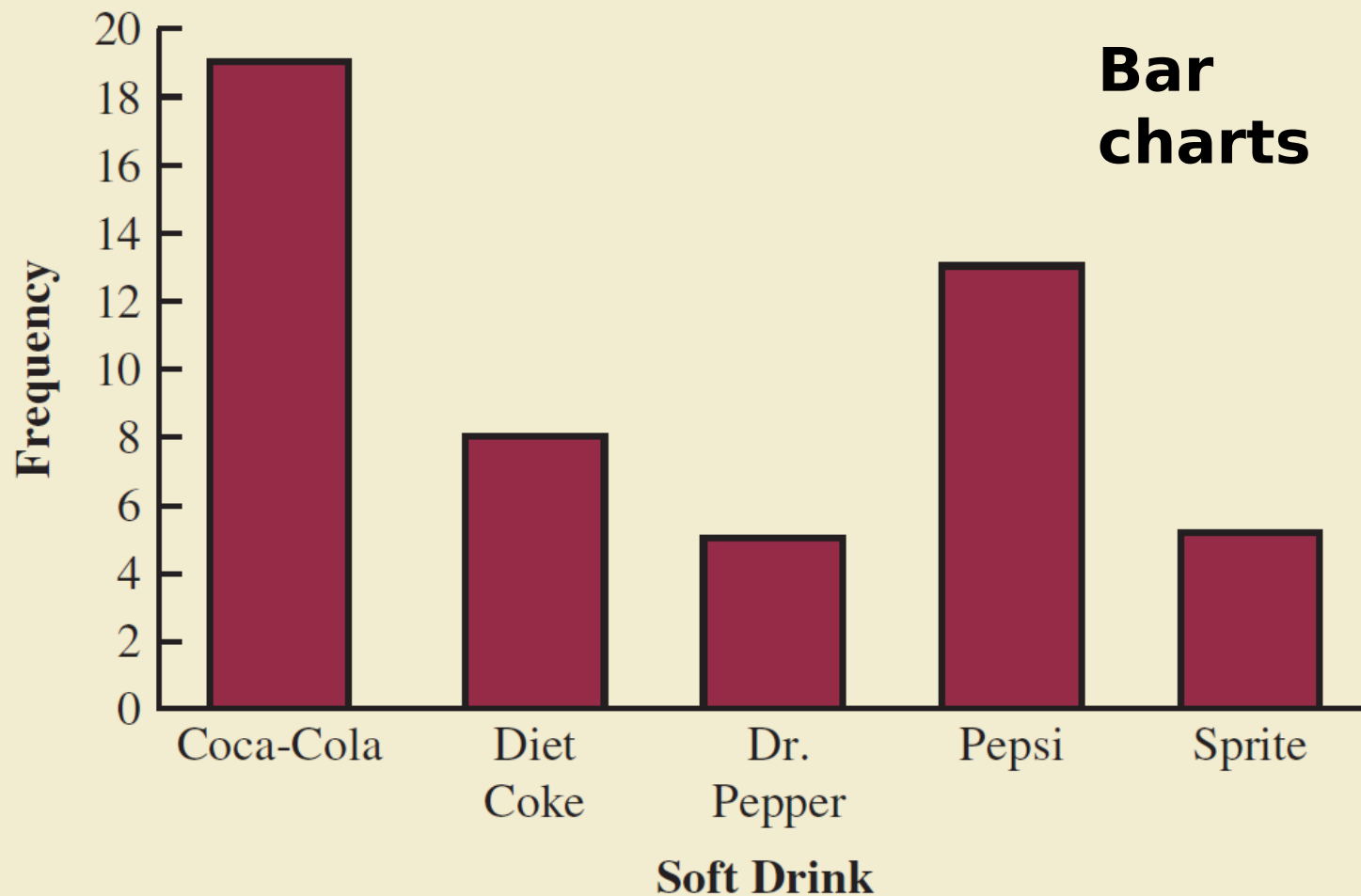
Soft Drink	Frequency
Coca-Cola	19
Diet Coke	8
Dr. Pepper	5
Pepsi	13
Sprite	<u>5</u>
Total	50

---





# Biểu đồ phân phối tần số



# TẦN SỐ TƯƠNG ĐỐI

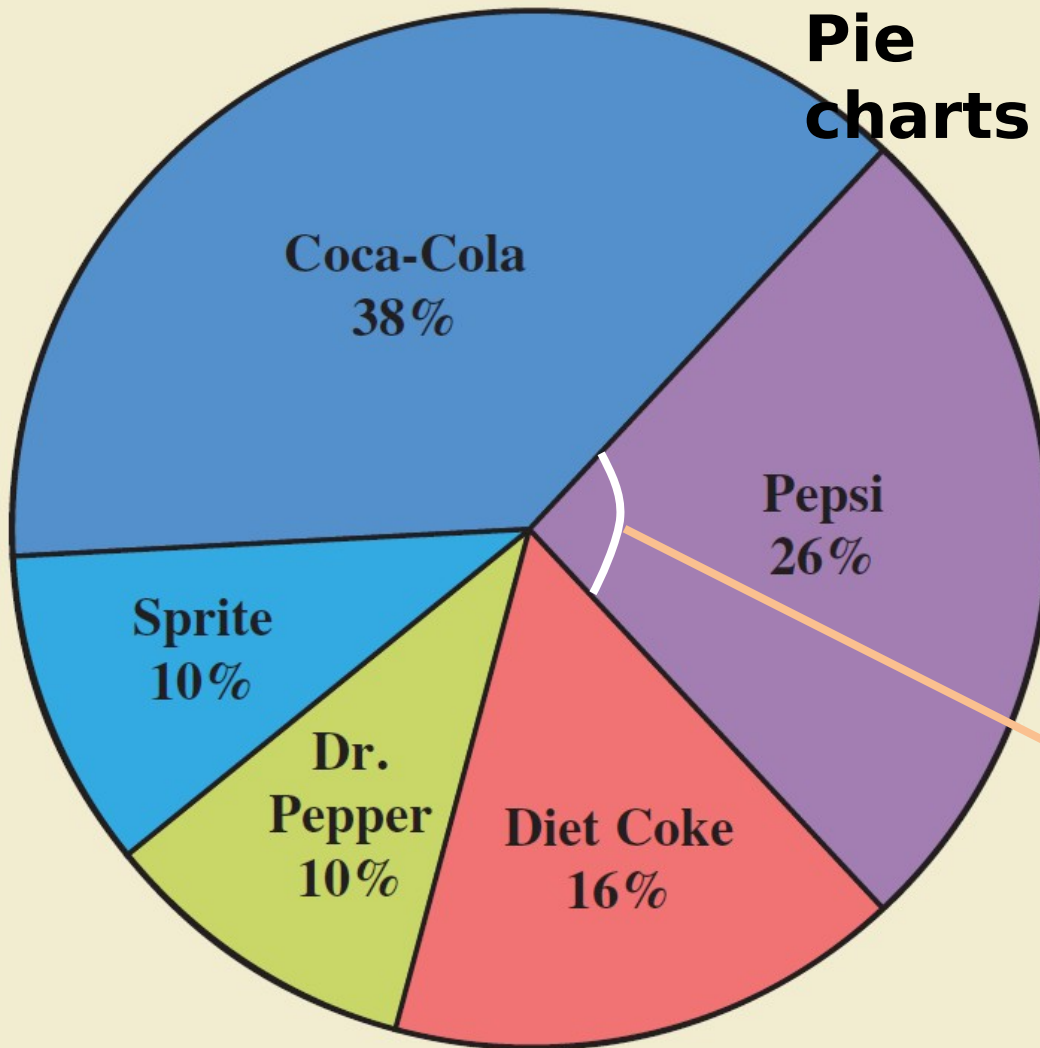
Soft Drink	Frequency	Relative Frequency
Coca-Cola	19	.38
Diet Coke	8	.16
Dr. Pepper	5	.10
Pepsi	13	.26
Sprite	5	.10
Total	50	1.00

$$\text{Tần số tương đối} = \frac{\text{Tần số } (f_i)}{\text{Tổng số } (n)}$$



# Biểu đồ phân phối tần số

**Pie charts**



$$0,26 \times 360^\circ = 93,6^\circ$$

# SẮP XẾP DỮ LIỆU ĐỊNH LƯỢNG

**TABLE 2.1** Recent Sale Price of Houses in Mission Viejo, CA, for June 2010 (data in \$1,000s)

\$430	670	530	521	669	445
520	417	525	350	660	412
460	533	430	399	702	735
475	525	330	560	540	537
670	538	575	440	460	630
521	370	555	425	588	430

Dữ liệu được chia làm k lớp với khoảng cách (h) đều nhau theo công thức:

$$k = \sqrt[3]{2n} \quad h = \frac{\text{Giá trị lớn nhất} - \text{Giá trị nhỏ nhất}}{k}$$



# NHẬN XÉT

---

1. Thường các lớp được chọn từ 5 đến 20 lớp.
2. Lớp (400 – 500) được hiểu là  $400 \leq X < 500$
3. Các lớp phải rời nhau, mỗi giá trị chỉ nằm ở một lớp duy nhất.
4. Tất cả các lớp phải bao gồm hết các giá trị của mẫu số liệu.
5. Không có lớp rỗng.



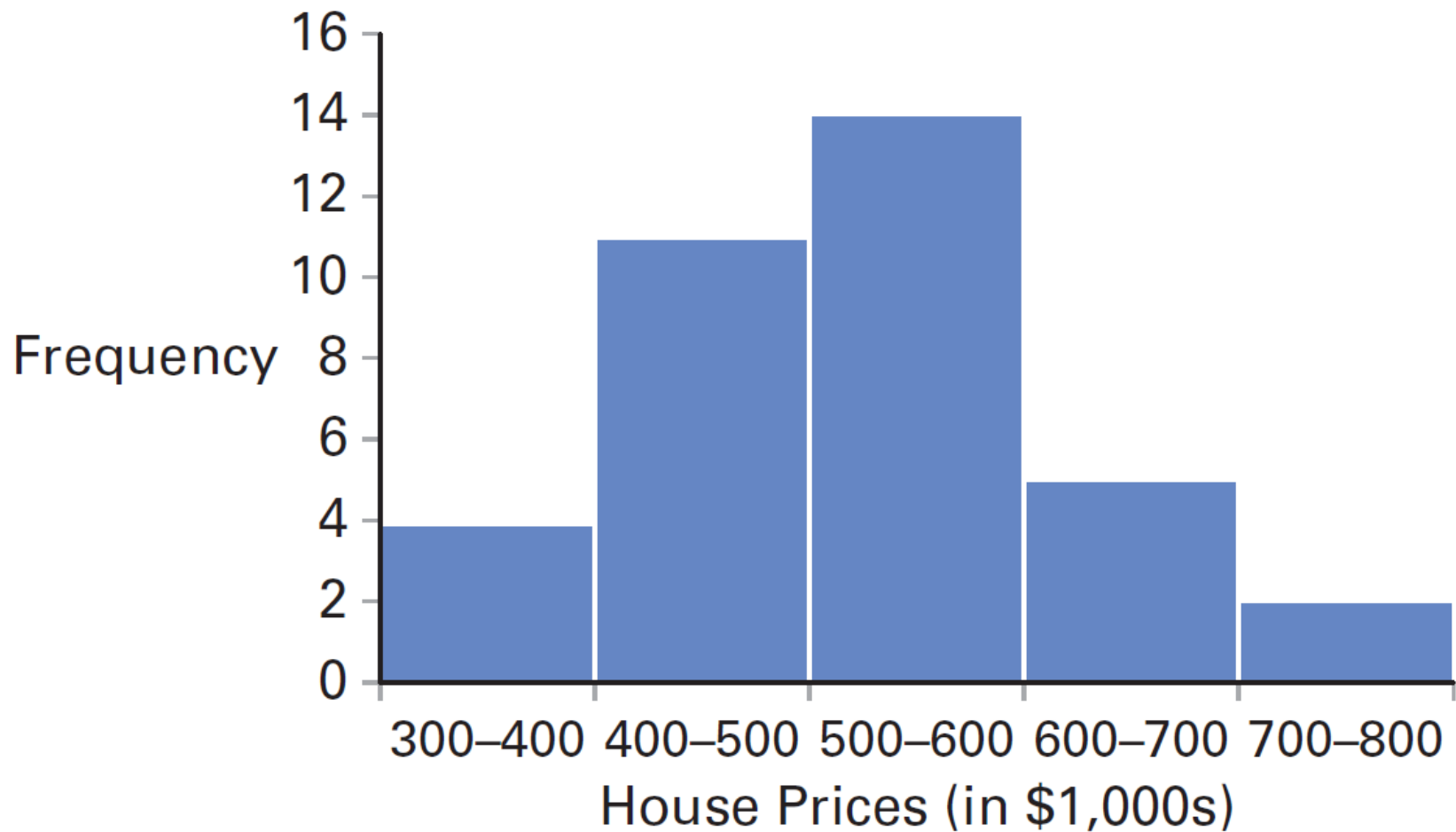


# BẢNG PHÂN PHỐI TẦN SỐ

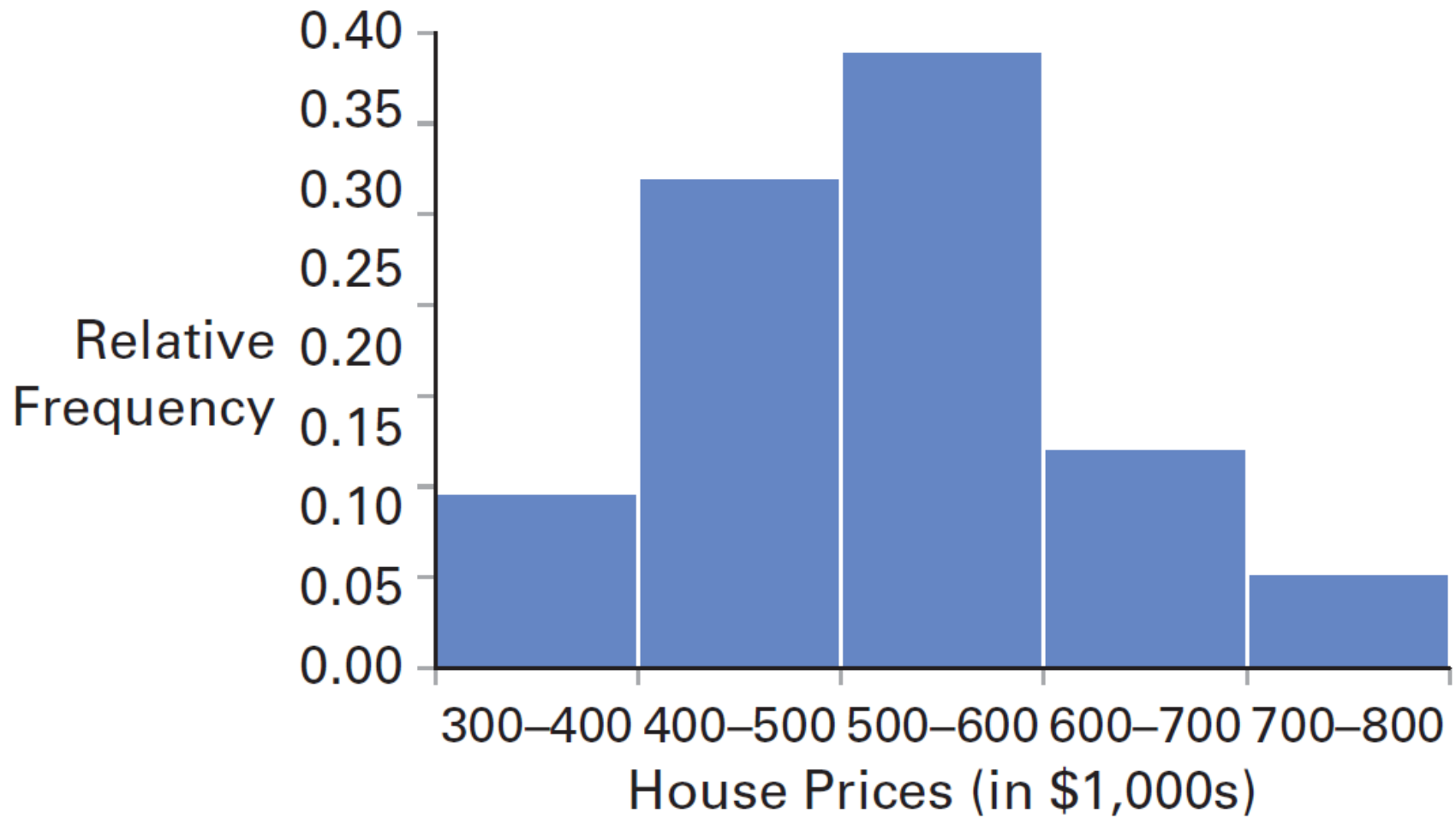
Class (in \$1000s)	Frequency
300 up to 400	4
400 up to 500	11
500 up to 600	14
600 up to 700	5
700 up to 800	2
	Total = 36



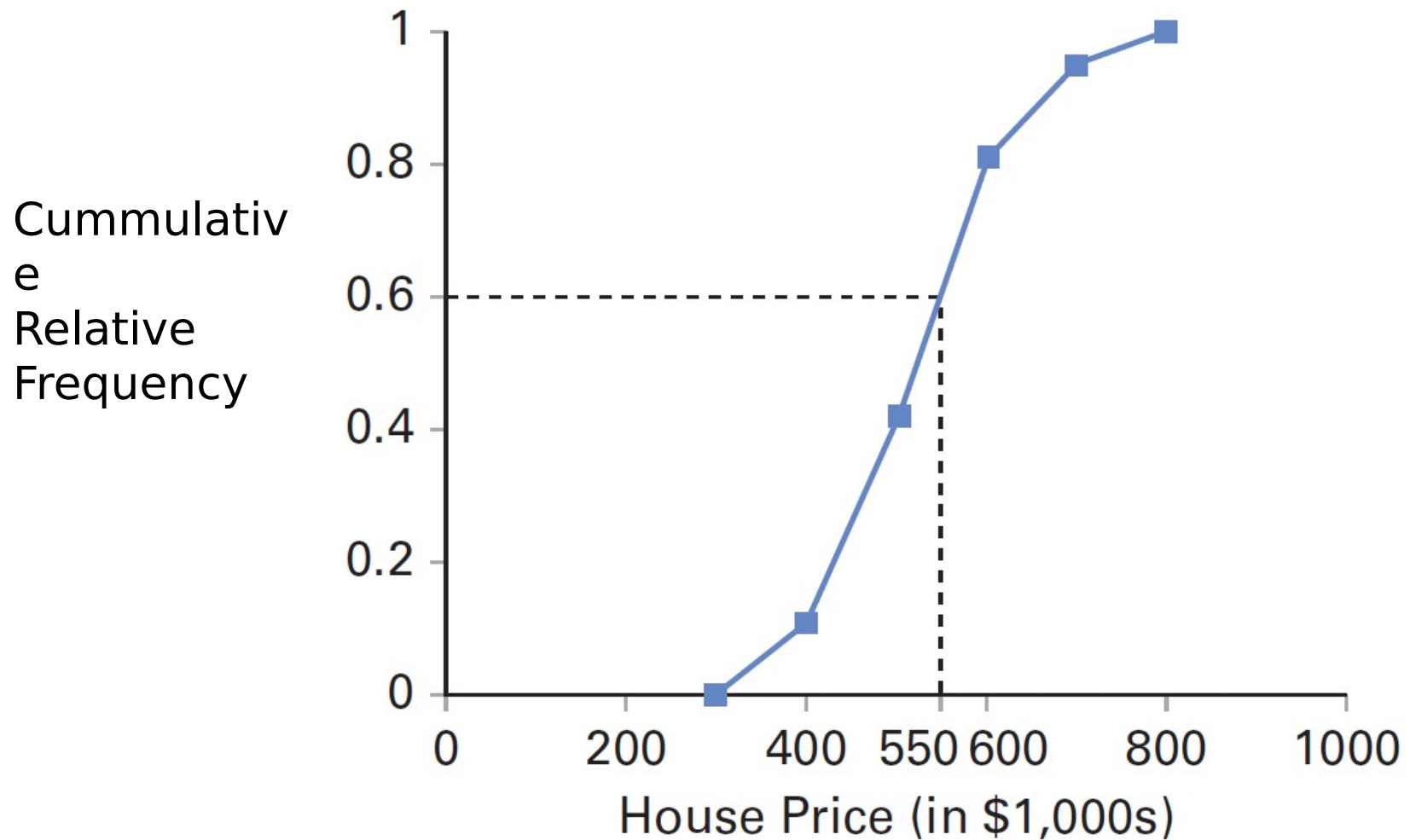
# BIỂU ĐỒ PHÂN PHỐI TẦN SỐ



# BIỂU ĐỒ PHÂN PHỐI TẦN SỐ TƯƠNG ĐỐI



# BIỂU ĐỒ PHÂN PHỐI TẦN SỐ TƯƠNG ĐỐI TÍCH LŨY



# PHƯƠNG PHÁP NHÁNH - LÁ (STEM & LEAF)

---



**John W. Tukey**  
**1915-2000**

Mỗi giá trị của tập dữ liệu được tách làm 2 phần: phần nhánh và phần lá.

Phần nhánh: chứa một hay nhiều chữ số đầu tiên.

Phần lá: chứa chữ số cuối cùng.





# VÍ DỤ

**TABLE 2.14** Wealthiest People in the World, 2010

Name	Age	Name	Age
Carlos Slim Helu	70	Li Ka-shing	81
William Gates III	54	Jim Walton	62
Warren Buffet	79	Alice Walton	60
Mukesh Ambani	52	Liliane Bettencourt	87
Lakshmi Mittal	59	S. Robson Walton	66
Lawrence Ellison	65	Prince Alwaleed Alsaud	54
Bernard Arnault	61	David Thomson	52
Eike Batista	53	Michael Otto	66
Amancio Ortega	74	Lee Shau Kee	82
Karl Albrecht	90	Michael Bloomberg	68
Ingvar Kamprad	83	Sergey Brin	36
Christy Walton	55	Charles Koch	74
Stefan Persson	62		

Reprinted by permission of Forbes Media LLC © 2011.

# VÍ DỤ

Stem	Leaf
3	6
4	
5	2 2 3 4 4 5 9
6	0 1 2 2 5 6 6 8
7	0 4 4 9
8	1 2 3 7
9	0



# VÍ DỤ

Stem	Leaf
3	6
4	
5	2 2 3 4 4 5 9
6	0 1 2 2 5 6 6 8
7	0 4 4 9
8	1 2 3 7
9	0

